



HAL
open science

ALEXIA : Un environnement d'aide à l'apprentissage lexical du françaislangue seconde

Thierry Chanier, Fabrice Issac, Christophe Fouqueré

► **To cite this version:**

Thierry Chanier, Fabrice Issac, Christophe Fouqueré. ALEXIA : Un environnement d'aide à l'apprentissage lexical du françaislangue seconde. Fiala P., Lafon P., Piguët M-P. La locution : entre syntaxe et pragmatique, INALF-Klincksieck, pp.105-118, 1997, <http://www.klincksieck.com/accueil.html>. edutice-00000631

HAL Id: edutice-00000631

<https://edutice.hal.science/edutice-00000631>

Submitted on 6 Oct 2004

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ALEXIA : Un environnement d'aide à l'apprentissage lexical du français langue seconde

Thierry Chanier, Fabrice Issac, Christophe Fouqueré. ALEXIA : Un environnement d'aide à l'apprentissage lexical du français langue seconde. Fiala P., Lafon P., Piguet M-P. *La locution : entre syntaxe et pragmatique*, INALF-Klincksieck, pp.105-118, 1997

Thierry CHANIER

*LRL - Université Clermont 2
34 avenue Carnot
F63037 Clermont-Ferrand*

Christophe FOUQUERÉ, Fabrice ISSAC

*LIPN - URA 1507, Université Paris-Nord
Av J.B. Clément
F93430 Villetaneuse*

RESUME : L'acquisition lexicale, et plus précisément l'acquisition des expressions lexicales, occupe une place importante en apprentissage de langue seconde. Nous présentons, dans cet article, un système d'apprentissage appelé ALEXIA. Celui-ci propose une aide lexicale personnalisée en production et en compréhension. Inconsidéré à la fois les stratégies d'accès lexicaux couramment employées pour modéliser la base de données lexicales, et celles employées par l'utilisateur. Le système propose également des activités lexicales pour l'aide à la rétention basées sur le lexique étudié par l'apprenant lors de ses travaux de compréhension ou/et de production. Encre qui concerne notre base de données lexicales, de nombreuses informations linguistiques sont disponibles. De plus, une aide à la correction des erreurs lexicales en production est accessible qui permet à la fois d'indiquer qu'un mot est mal employé, et de proposer expressions ou collocations mieux adaptées à ce qui est supposé être produit.

1. Introduction

L'acquisition lexicale, et plus précisément l'acquisition des expressions lexicales, occupe une place importante en apprentissage de langue seconde (L2). En effet, le fait qu'un apprenant ne maîtrise pas un niveau de vocabulaire suffisant est considéré comme un facteur fortement préjudiciable à la communication. Les études en linguistique théorique et appliquée, ainsi qu'en psycholinguistique montrent que le mot est une unité linguistique complexe porteuse de connaissances morphologiques, syntaxiques, sémantiques et même pragmatiques.

De plus, le choix d'un mot dans la production d'un énoncé peut imposer une structure sur tout l'énoncé. Connaître un mot c'est ainsi tout à la fois savoir dans quel contexte il est utilisé à l'oral ou à l'écrit (fréquence d'utilisation, collocations associées), appréhender les limitations

de son usage selon les variations de fonctions ou de situations, connaître ses comportements morphologiques, syntaxiques et sémantiques, et enfin pouvoir l'intégrer dans son propre lexique mental et établir les relations nécessaires avec les autres mots dans le réseau associé [Richards 85].

Rappelons que le taux de l'acquisition du vocabulaire en langue seconde est étonnamment faible. Certains chercheurs estiment ainsi que l'acquisition de 2000 mots en 5 ans est au-delà de la capacité de beaucoup d'apprenants[1]. L'apprentissage d'une L2 est donc délicat à plus d'un titre : temps d'apprentissage court, non intensif, rendement faible, donc mauvaise utilisation des stratégies adoptées en langue native. Il y a donc un défi majeur à relever pour le domaine de l'apprentissage de langues assisté par ordinateur (ALAO).

Un certain nombre de systèmes d'apprentissage se sont orientés vers la mise à disposition des apprenants de ressources lexicales "traditionnelles" (dictionnaire électronique organisé alphabétiquement, concordanceurs et bases de textes plus ou moins homogènes) autour desquelles ils construisent des exercices lexicaux et surtout des activités où l'apprenant peut collecter et organiser ses informations lexicales. Pour notre part, compte tenu de ce que l'on connaît de l'organisation du lexique mental [Aitchison 87] et la définition que l'on donne de l'acquisition d'un mot (cf. ci-dessus), il nous est apparu important d'organiser ces activités lexicales autour d'une base de données "active" contenant toutes les informations nécessaires (morphologiques, syntaxiques, sémantiques et pragmatiques). Par "active", nous entendons le fait que le système sait mettre en œuvre ses connaissances pour :

- en aide à la compréhension, offrir des accès multiples à l'apprenant, visualiser sous formes de réseaux multiples les associations entre mots,
- en aide à la production, diagnostiquer les énoncés de l'apprenant de façon à savoir non seulement corriger certaines erreurs, mais également offrir des paraphrases.

Enfin, nous avons également décidé de ne pas restreindre le lexique aux mots simples, mais d'y inclure également les collocations et expressions semi-figées (dont les expressions idiomatiques). Ces expressions sont très peu considérées en ALAO. Pourtant ces unités linguistiques sont numériquement beaucoup plus importantes dans les dictionnaires électroniques de langues telles que le français ou l'anglais que les entrées lexicales simples[2]. Elles sont d'un usage fréquent par les natifs, mais sont difficiles à acquérir par les apprenants non-natifs. Aux yeux des didacticiens en L2, les justifications pour les enseigner sont multiples [Nattinger 88] : ces expressions fournissent un matériel de base pour l'analyse et la segmentation du discours, leur maîtrise doit permettre à l'apprenant de ne pas violer certaines restrictions lexicales, lui éviter de commettre des erreurs de registre dans la production de son discours, elles peuvent faciliter ses productions écrites ou orales puisqu'elles lui évitent de porter une attention soutenue à chaque mot pris individuellement, enfin elles autorisent l'apprenant à porter son attention sur des structures plus larges du discours et sur les aspects sociaux de l'interaction.

Un certain nombre d'environnements d'aide à l'apprentissage lexical cherchent à remédier à cet état de fait. L'aide apportée est toutefois orientée production, ou orientée compréhension, mais pas aux deux [Fontana & al 93, Goodfellow 93]. Cette aide sera parfois très "imposée" : la démarche de l'apprenant n'y est pas prise en compte. Si les stratégies cognitives d'accès et de rétention de vocabulaire sont considérées, elles le seront lors de la construction de la base de données du système et non pas lors de son utilisation [Agirre & al 90].

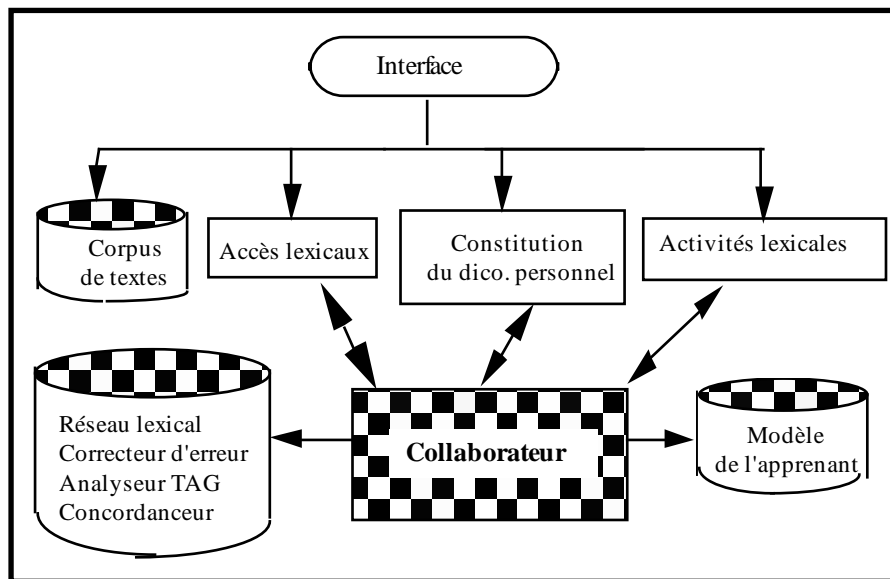
Notre système adopte une approche différente. En ce qui concerne l'utilisateur, il propose une aide lexicale personnalisée en production et en compréhension. Il considère à la fois les stratégies d'accès lexicaux couramment employées pour modéliser la base de données lexicales, et celles employées par l'utilisateur. Celui-ci a la possibilité de déconstruire une base de données lexicales suivant ses propres critères. Le système propose également des activités lexicales pour l'aide à la rétention basées sur le lexique étudié par l'apprenant lors de ses travaux de compréhension ou/et de production. En ce qui concerne notre base de données lexicales, de nombreuses informations linguistiques sont disponibles. De plus, une aide à la correction des erreurs lexicales en production est accessible qui permet à la fois d'indiquer qu'un mot est mal employé, et de proposer expressions ou collocations mieux adaptées à ce qui est supposé être produit. Enfin, deux types de représentations lexicales sont adoptés permettant à l'utilisateur de visualiser les renseignements demandés : réseau et arbre. La modélisation des données lexicales et des accès lexicaux devient dès lors le problème essentiel d'un tel environnement.

La première partie présente succinctement notre système ALEXIA. Les deuxième, troisième et quatrième parties exposent le formalisme des données accessibles et les traitements afférents. La dernière partie expose les différents aspects de l'accès et de l'apprentissage lexical dans notre système et conclue sur les perspectives offertes par une telle approche.

2. ALEXIA : Présentation

Ce système s'adresse à des apprenants avancés en français langue seconde. Il a pour but d'aider celui-ci, de manière personnalisée, à la mémorisation et la structuration de l'information lexicale, et plus spécifiquement en ce qui concerne les expressions et collocations.

Le système ALEXIA est composé de différents modules (cf. Figure 1) : accès lexicaux, activités lexicales, modèle de l'apprenant, et suppose deux bases de données : un corpus de textes et un réseau lexical. Nous allons décrire ce système au travers des scénarios types d'utilisation par l'apprenant. Notons d'ores et déjà que l'ensemble des tâches effectuées par l'apprenant sont enregistrées dans le module "modèle de l'apprenant". Celui-ci regroupe à la fois les stratégies employées au niveau des accès lexicaux et la création d'une base lexicale personnelle. Les données de ce module sont réutilisées afin de proposer à l'apprenant des *activités lexicales* dont le but est de l'aider à retenir les nouveaux éléments lexicaux rencontrés et étudiés. La base lexicale est décrite plus amplement en section 4. Celle-ci est implémentée sous forme de réseau. En effet, les apprenants effectuent fréquemment des associations entremets selon différents critères (même catégorie syntaxique, même champ sémantique, etc.) améliorant ainsi sensiblement la rétention des informations ; remarquons que nous appliquons en cela les théories actuelles concernant l'organisation du lexique mental [Aitchison 87, Singleton & Little 91].



Architecture simplifiée d'ALEXIA

Figure 1

Deux scénarios d'utilisation sont proposés l'utilisateur. Le scénario de compréhension permet l'utilisateur de consolider son vocabulaire, le scénario en production permet de le valider.

2.1. Scénario type d'utilisation en compréhension

La 'compréhension' correspond à deux tâches complémentaires : compréhension proprement dite d'un vocabulaire de textes, mémorisation de ce vocabulaire (i.e. augmentation du lexique de l'apprenant). L'apprenant travaille sur un des textes proposés dans le corpus de textes. Ce corpus est décrit plus avant en section 3, l'utilisateur a la possibilité de spécifier un type de texte, i.e. de spécifier le domaine lexical correspondant à son centre d'intérêt. La démarche classique consiste alors à relever tous les nouveaux mots et expressions qu'il n'a jamais rencontrés ou tous ceux qu'il a déjà rencontrés mais dont la compréhension lui semble parcellaire.

Le module d'accès lexicaux permet à l'apprenant de demander différents types d'informations linguistiques sur les items lexicaux qu'il a relevés : définition, exemples d'utilisation, synonymes, antonymes, dérivés, ... A ce stade, il peut également faire appel au système pour qu'il lui indique les expressions ou collocations correspondant à un sens particulier. Il peut également accéder aux représentations graphiques de portions du réseau sémantique, et aux représentations syntaxiques (i.e. structure syntaxique dans laquelle le mot peut être utilisé) des mots et expressions lexicales qu'il étudie. Ces informations lui apportent en outre des informations sur les variations lexicales et les transformations syntaxiques possibles.

Après cette première phase de travail où l'apprenant a demandé un certain nombre d'aides lexicales, l'utilisateur complète sa propre base de données en utilisant une des méthodes ci-dessous :

* Annotation :

Il s'agit dans ce cas de "prendre des notes" sur chaque mot ou expression, sur lesquels il a eu besoin d'avoir des renseignements. Parmi ceux-ci, notons : la traduction dans sa langue

maternelle, certaines des informations extraites du réseau lors de ces précédents accès lexicaux.

* Association :

L'association personnelle de mots et/ou expressions entre eux est une démarche féconde en apprentissage. Elle permet ainsi à l'utilisateur l'appropriation effective des renseignements obtenus

Exemple : création d'un groupe "tromperie", dans lequel on trouve les éléments "mener en bateau", "tromper", "berner".

* Graphisme :

L'utilisateur peut enfin "dessiner son propre réseau lexical", les nœuds représentant les mots qu'il a étudiés et les liens traduisant, par exemple, les associations qu'il a faites auparavant, ou des portions du réseau de base.

Les différents accès et travaux effectués par l'apprenant sont enregistrés et forment le modèle de l'apprenant.

2.2. Le scénario type d'utilisation en production

L'apprenant peut soit faire le résumé d'un texte choisi dans le corpus de base soit produire un texte sur un thème précis. La sélection d'un texte ou d'un thème est un prérequis essentiel à l'analyseur grammatical (cf. section 5). Celui-ci propose une correction des erreurs aux niveaux grammatical et lexical. Le système peut alors fournir à l'apprenant une formulation mieux adaptée de ce qu'il a écrit, en particulier, lui indiquer les expressions qui correspondent plus exactement à ce qu'il a voulu exprimer. On retrouve dans ce cas les mêmes fonctionnalités des accès lexicaux qu'au niveau de la compréhension. L'aide à la production est ainsi à la fois dynamique et statique. L'aide dynamique correspond à une utilisation orientée des bases de données lexicales (lexique personnel ou fourni par le système). L'aide statique correspond à l'amélioration syntaxique et lexicale proposée par l'analyseur.

3. Corpus et Données lexicales

Le choix de décrire finement les connaissances lexicales nécessite de fixer un champ sémantique. De façon à ne pas disperser nos recherches, nous nous sommes fixés sur les champs travail, emploi et chômage. Puisque l'expertise doit se référer à la façon dont la langue est réellement utilisée, nous avons construit un corpus électronique de textes de français courant, à partir duquel ont ensuite été menées des études linguistiques, qui serviront à la création de la base de données lexicales active.

Corpus

Afin de construire notre expertise lexicale sur le champ sémantique de la recherche d'un emploi, nous avons commencé à construire un corpus électronique, qui compte aujourd'hui 250 000 mots, du français courant. Pour ce faire, nous avons sélectionné des articles de

journaux, revues, magazines de toutes natures : presse économique ou sociale, quotidiens "populaires" ou non, périodiques édités par les associations de chômeurs, etc. Ce corpus, composé de textes écrits, contient néanmoins des interviews.

Les raisons de la constitution de ce corpus sont multiples, il doit : comme nous l'avons dit servir à construire un dictionnaire des termes du domaine ; fournir des exemples d'utilisation de ces termes en contexte, en associant item lexical et phrases dans lesquelles il apparaît ; être une base de textes pour la constitution d'activités lexicales d'aide à l'acquisition lexicale en L2.

Extraction des mots et expressions du domaine

Pour exploiter les données de ce corpus, nous avons ensuite élaboré une procédure d'extraction automatique des items lexicaux (mots simples, collocations et expressions terminologiques) caractéristiques du domaine [Rabefitia 94]. Pour ce faire, nous avons choisi les méthodes statistiques et probabilistes. Pour extraire les mots de base nous avons calculé l'ordre lexicométrique en relevant la liste de chaque forme du corpus associée à sa fréquence, par ordre décroissant. Les premières formes de la liste correspondaient, bien entendu, aux mots grammaticaux, mais on y trouvait également des mots lexicaux, dont l'utilisation exceptionnellement fréquente était conditionnée par la thématique du corpus (pour les substantifs "travail", "embauche", "emploi", etc.). Ces mots lexicaux, associés à quelques verbes support ("faire", "mettre", etc.) souvent employés dans les expressions, ont constitué les mots de base.

Cette liste des mots de base a servi de référence pour l'extraction des expressions, dont un des constituants devaient appartenir à cette liste. Cette façon de procéder élimine certaines expressions, comme les expressions à caractère métaphorique (puisque'il se peut qu'aucun des mots d'une telle expression n'appartienne au champ lexical étudié), mais présente l'avantage de réduire le bruit (les structures ne caractérisant pas le domaine, mais la langue française en général). Ces cooccurrences ont été construites à partir des lemmes et non des formes fléchies, de façon à observer les variations syntaxiques, le degré de figement, de ces expressions.

A partir de cette liste de cooccurrences et de l'ordre lexicométrique du corpus, nous avons pu appliquer un modèle probabiliste. Nous nous sommes centrés sur les lemmes des mots de base et avons éliminé les mots grammaticaux présents dans les associations. Ayant obtenu les probabilités pour chaque collocation et fixé des seuils d'acceptabilité, nous sommes intervenus une seconde fois manuellement pour éliminer (10 à 15% des collocations) les collocations qui ne présentent pas d'intérêt du point de vue de leur utilisation dans le champ lexical de la recherche d'emploi. Ces deux interventions manuelles permettent de réduire la récolte trop large résultant d'approches purement statistiques et probabilistes. Ensuite chaque expression retenue a été indexée avec les phrases dans lesquelles elles apparaissaient, de façon à en étudier les variations syntaxiques et à servir d'illustration pour les apprenants.

Études linguistiques

150 unités lexicales ont été finalement retenues et décrites, une à une, à partir d'un patron très proche de celui utilisé dans la théorie sens-texte de Mel'cuk [Mel'cuk 92]. Le dictionnaire explicatif et combinatoire offre sans doute la description (partielle) la plus fine du français, description en partie opératoire et orientée vers la génération et la paraphrase :

1) Les informations de bases

- La désambiguïsation du sens. Pour une unité lexicale donnée plusieurs sens sont possibles. Une entrée lexicale correspond toutefois à un seul sens, les différentes acceptions d'un mot ou expression étant indexées.

Exemple : travailler

1. X Avoir un métier
2. X Faire Y avec un certain effort
3. X Modifier Y par une action suivie

- Les définitions. Elles sont soit calculées à partir des informations disponibles dans le réseau, soit construites à partir de primitives.

Exemple :

1. `bosser' a pour définition : Plus (travailler) \Leftrightarrow travailler dur
2. `travailler' : faire une chose avec un certain effort

- Les exemples d'utilisation en contexte. Un certain nombre d'exemples d'utilisation en contexte du lexème d'entrée sont extraits des textes du corpus comme informations associées à l'entrée lexicale de ce lexème.

- Les registres de langue. Quatre registres ont été retenus : soigné, courant, familier, grossier.

2) Les informations syntaxiques

On indique, catégorie syntaxique, genre et nombre pour chaque entrée lexicale. On donne également des renseignements concernant les constructions syntaxiques et un certain nombre de transformations syntaxiques possibles (passivation, pronominalisation, question, introductions de modificateurs, variations lexicales, ...) pour l'entrée lexicale considérée (cf. aussi section 4).

3) Les informations du réseau

Le réseau est formé de nœuds représentant les entrées lexicales, et de liens représentant différentes associations. Les liens sémantiques calculables. Chaque item lexical est relié à différents autres par plusieurs types de liens définis ci-après. Les entrées lexicales étant gérées par le sens, les expressions et mots peuvent être reliés entre eux, ainsi que les expressions entre elles.

Les expressions sont liées :

* par le sens : expressions de même sens ou ayant des liens de types sémantiques entre elles,

* par le mot : des expressions comportant un mot commun sont également reliées entre elles par ce mot.

La liste des liens retenus, et de leur fonction, pour la modélisation du réseau se divise en quatre groupes :

* Type sémantique : synonymie et antonymie, exactes, plus larges, plus étroites, ou intersection de sens.

* Types dérivés, actantiels et circonstanciels.

* Fonctions : trois fonctions ont été retenues pour leur caractère très général leur permettant de s'appliquer dans de nombreuses situations.

Exemple : Oper1(travail): dénicher [un travail]

* Composants de sens : Moins, Plus, Très/Intense, Bon, ...

4. Analyse

Nous avons choisi d'utiliser le formalisme des grammaires d'arbres adjoints [Abeillé 90] pour l'analyse de phrases dans ALEXIA. Ce formalisme, qui se situe dans la lignée des grammaires d'unification, est basé non sur des règles de réécriture mais sur des structures d'arbres élémentaires. L'analyse repose sur une opération spécifique : l'adjonction, et comporte une contrainte importante : la lexicalisation des informations linguistiques. Tout arbre élémentaire doit ainsi avoir au niveau de ses feuilles au moins une tête lexicale. Cette lexicalisation permet de rendre compte correctement des processus linguistiques en jeu dans les expressions lexicales, par la création des arbres élémentaires correspondant.

L'analyse proprement dite se compose de deux phases : initialisation et regroupement, décrite dans [Issac 94]. Durant la phase d'initialisation, on parcourt la grammaire afin de créer une sous-grammaire minimale. Puis on détermine pour chaque arbre les différentes positions d'adjonction possibles. La phase de regroupement consiste à adjoindre ou substituer selon les cas les arbres correspondant aux sous-chaînes contiguës de la chaîne à analyser. L'analyse est essentiellement montante afin d'obtenir, en cas de phrase incorrecte, le plus d'informations partielles. En cas d'échec, les arbres partiels sont présentés simplifiés à l'apprenant afin que celui-ci puisse rectifier la construction de sa phrase. L'apprenant a ainsi simultanément et la possibilité de corriger sa production, et de comprendre les contraintes inhérentes aux expressions ou mots qu'il a choisis.

5. Stratégies cognitives et apprentissage

Compte tenu des objectifs de l'environnement ALEXIA, nous avons besoin d'évaluer :

- les stratégies adoptées par les apprenants dans des tâches de compréhension d'un texte, celles adoptées dans des tâches de production,
- les moyens que les apprenants utilisent pour construire leur lexique personnel,
- dans quelle mesure les différents types d'accès lexical et les ressources lexicales disponibles aident à la compréhension, à la production et à la rétention de vocabulaire.

Les travaux en psycholinguistique ou en linguistique appliquée n'apportent pas de réponses à ces questions pour les raisons suivantes. Les expérimentations menées en psycholinguistique sur le lexique (cf. par exemple [Taylor 90]) consistent généralement à inférer la structure du lexique mental à partir de tâches de reconnaissances, conduites en temps très limité, de mots ou segments de phrases dans des contextes très restreints. Les réponses des sujets sont limitées à des sélections de choix prédéfinis. Même si nous prenons en considération les résultats de ces expériences, la méthodologie expérimentale utilisée ne permet pratiquement jamais de mesurer les performances des sujets en situation de production lexicale. Le contexte trop restreint ne prend pas en compte la combinatoire linguistique rencontrée dans les textes et oblitère donc les choix de hauts niveaux que doivent accomplir ces sujets. Enfin la vitesse de réponse des sujets aux stimuli ne donne aucune information sur les stratégies de choix mises en œuvre sur des durées plus longues, c'est à dire lorsque les sujets ont à intégrer un ensemble d'informations provenant du contexte textuel et à opérer une sélection sur ces informations.

Quant aux expériences, en nombre limité, menées par les chercheurs de linguistique appliquée s'intéressant au lexique, elles ont pour support unique les dictionnaires papiers [Bogaards 91] et s'intéressent à des tâches de traduction. Les résultats de ces expériences sont difficilement transposables puisque les tâches données aux sujets ne sont pas celles généralement rencontrées en situation de communication en langue seconde. De plus, ces expériences ignorent les environnements informatiques, ainsi que toutes les stratégies spécifiques induites par ces environnements sur les apprenants.

Nous préparons donc une expérimentation qui constitue une évaluation a priori des hypothèses sous-jacentes dans ALEXIA et pourrait ainsi nous amener à réorienter le développement de notre environnement informatique si nécessaire. Souvent, pour les environnements informatiques, ce type d'évaluation est réalisé a posteriori (après le développement du logiciel). Il sert alors généralement plus à justifier le travail réalisé qu'à le modifier.

Expérimentation

L'environnement ALEXIA étant basé sur un réseau lexical, il est nécessaire de pouvoir mener une expérimentation sur ce type de support. Il existe déjà deux dictionnaires électroniques qui permettent une recherche lexicale fondée sur les liens sémantiques entre items lexicaux : l'un en anglais, WORDNET, et l'autre en français, DICOLOGIQUE [Dutoit 92]. A notre connaissance il n'y a pas eu d'expériences, répondant à nos objectifs, menées sur ces deux environnements. Notre système est orienté vers l'apprentissage lexical du français, justifiant ainsi le choix de DICOLOGIQUE comme cadre d'expérimentation. Mener une expérimentation doit nous permettre d'observer comment les sujets exploitent ce type de réseau au cours des tâches qu'ils ont à réaliser et de répondre aux questions suivantes :

- Comment les apprenants procèdent-ils pour accéder à un item lexical? (Accès)

- Quelles stratégies emploient-ils pour comprendre le sens d'un mot nouveau? (Compréhension)
- Quelles sont les stratégies employées lors de l'utilisation d'un nouveau mot ? (Production)
- Quelles sont les méthodes employées par les apprenants pour retenir le vocabulaire, l'utilisation d'un environnement informatique basé sur un réseau lexical favorise-t-elle cette rétention? (Apprentissage)
- Un réseau lexical permet-il une meilleure approche des tâches de production et de compréhension à effectuer, quels sont les avantages et les désavantages de la représentation adoptée: est-elle mieux adaptée aux stratégies des apprenants et correspond-elle mieux à leurs représentations mentales ? (Apports d'informations pour la construction de notre réseau lexical).
- Quelles sont les informations les plus demandées dans une recherche lexicale (synonymies, antonymie, structures syntaxiques, exemples, définitions...) ?

L'expérimentation se déroulera auprès d'un public d'apprenants de français langue étrangère de niveaux différents en formation. Les sujets seront donc en phase d'acquisition lexicale soutenue, mais auront sur le plan des champs sémantiques les mêmes connaissances qu'un sujet natif du français. Le fait de prendre des sujets de niveaux différents nous permettra de comparer leurs stratégies utilisées, de façon à pouvoir ensuite, dans ALEXIA, introduire un niveau réflexif permettant de dialoguer avec chaque apprenant sur sa façon de travailler.

Lexique personnel et activités lexicales

Le premier principe pédagogique défendu par Goodfellow [Goodfellow 93] est l'assistance à la construction d'un dictionnaire propre à l'apprenant. Trois critères, interdépendants, justifient l'appartenance d'un item lexical à ce dictionnaire personnalisé : une expression pourra avoir été abordée en compréhension, en production, ou introduite dans des regroupements d'expressions opérés directement par l'apprenant. Tout comme la présentation d'informations lexicales sous forme de réseaux favorise l'acquisition, Goodfellow a montré qu'un environnement informatique devait permettre à l'utilisateur d'effectuer ses regroupements personnels. La question reste ouverte de savoir si les types de liens entre les items lexicaux doivent être laissés au libre arbitre de l'utilisateur ou, au contraire, être choisis parmi ceux effectivement utilisés par la base lexicale. Le statut d'un item lexical du dictionnaire peut donc être différent suivant le critère qui a présidé à son mode d'introduction, et suivant qu'il ait, ou non, été validé par des activités lexicales appropriées.

MacWhinney suggère, à partir de ces travaux sur le Modèle de Compétition, quelques principes pédagogiques suivant les stades d'acquisition lexicale d'un apprenant (nous omettons le stade de restructuration phonologique et de transfert initial) :

- l'apprentissage par cœur d'items lexicaux joue un rôle important au début, mais devra ensuite s'effacer,
- lors de l'acquisition d'un nouvel item lexical, il est nécessaire d'explicitement sa structure syntaxique et casuelle et de le relier explicitement à d'autres items lexicaux. A ce stade, des erreurs dues à des transferts depuis la L1 devraient être corrigées explicitement,

- même si l'application de simples procédures de transfert continue à être source d'erreurs, il vaut mieux ne pas chercher à détecter, ni corriger de telles erreurs en production. La compréhension, essentielle à ce stade, doit être travaillée à partir de matériaux riches et difficiles,

- à un stade plus avancé, il est nécessaire de détecter et corriger les erreurs de façon à éviter l'apparition de phénomènes de fossilisation et d'aider à la restructuration fonctionnelle des connaissances de l'apprenant.

A partir de cette description générale, selon le niveau de l'apprenant, il est possible de construire une série d'activités lexicales traditionnelles en compréhension et production sur des mots ou expressions : reconstitution d'une partie de réseau, groupements de mots suivant différents critères, exercices à trous, paraphrase d'une expression, choix entre différentes paraphrases en contexte, détermination du sens d'un mot ou d'une expression à partir d'un texte dans lequel, il (elle) est introduite de façon redondante, retrouver le sens d'un mot ou d'une expression à partir d'informations fragmentaires (concept du "mot sur le bout de la langue"), etc. Ces activités nécessitent pour une grande part l'utilisation du corpus de textes, de la base de données lexicales et de l'analyseur TAG, lorsqu'il sera nécessaire de diagnostiquer les productions de l'apprenant.

Au travail linguistique proprement dit, il convient d'ajouter la possibilité d'un dialogue avec l'apprenant sur la tâche qu'il a à accomplir, c'est-à-dire de pouvoir discuter explicitement des stratégies qu'il emploie, en particulier pour l'accès lexical.

Cette construction du dictionnaire personnalisé et l'introduction d'activités acquisitionnelles correspondant aux recommandations des chercheurs en psycholinguistique et linguistique appliquée ne peut avoir lieu sans la mise en place de ressources linguistiques importantes, organisées en fonction des besoins des apprenants. Les ressources que nous avons mentionnées à la section précédente font, encore aujourd'hui, défaut aux systèmes lexicaux d'ALAO. C'est donc prioritairement sur ce point que portent nos efforts actuels.

Bibliographie

Abeillé A. [1990] : "Quand l'arbre ne cache pas la forêt, analyse du français à l'aide d'une grammaire d'arbres adjoints": in *T.A. Informations*, Vol. 31, ndeg.2, 1990, pp 51-70.

Agirre E., Arregi X., Artola X., Diaz de Ilarraza A., Evrard F., Sarasola K.[1990] : "Dictionnaire intelligent d'aide à la compréhension", Proceedings of *EURALEX'90*, 1990, pp 45-57.

Aitchison J. [1987]: *Words in the mind*. Oxford: Blackwell, 1987.

Bogaards P. [1991] : "A propos de l'usage du dictionnaire de langue étrangère", in *Cahiers de lexicologie*, ndeg.52, vol.1,1991, pp 131-152.

Dutoit D. [1992] : "A set theoretic approach to lexical semantics", Actes de *COLING 92*, Nantes, 1992.

- Fontana N.M., Caldeira S.M., Cristina M., De Oliveira F., Oliveira Jnr O. N.[1993] : "Computer Assisted Writing: Applications to English as a Foreign Language", in *Computed Assisted Language Learning*, vol 6, ndeg.2, 1993,pp 145-162.
- Goodfellow R. [1993] : "Call for vocabulary, requirements, theory and design", in *Computer Assisted Language Learning*, vol 6, ndeg. 2, 1993, pp99-122.
- Issac F. [1994] : "Un algorithme d'analyse pour les grammaires d'arbres adjoints", Colloque international sur les *Grammaires d'Arbres Adjoints*. Paris, 1994.
- Mel'cuk I. [1992] : "Introduction", in *Dictionnaire Explicatif et Combinatoire du français Contemporain. Recherche lexico-sémantiques III*. Les presses de l'université de Montréal, 1992, pp 9-58.
- Nattinger J. [1988] : "Some current trends in vocabulary teaching", in *Vocabulary and Language Teaching*. Carter R., McCarthy M. (Eds.) Longman,1988, pp 62-81.
- Rabefitia M. [1994]: *Extraction automatique des collocations et des expressions terminologiques*. Mémoire de DEA MIASH,Université Paris 4 et Paris 5, 1994.
- Richards J.C. [1985] : "Chapter 13 : Lexical knowledge and the teaching of vocabulary", in *The Context of Language Teaching*, Richards J.C. (Ed.),Cambridge University Press, 1985.
- Singleton D., Little D. [1991] : "Le lexique mental de l'apprenant d'une langue étrangère", in *Acquisition et enseignement/apprentissage des langues*, Colloque "Acquisition d'une langue étrangère: perspectives et recherches", Grenoble, 1991, pp 395-402.
- Taylor I. [1990] : *Psycholinguistics: Learning and Using Language*. Prentice Hall, 1990.

[1] En langue native, l'acquisition est de 3000mots par an durant la période scolaire.

[2] Les travaux au LADL et ailleurs ont montré leur forte proportion dans la langue (20000 expressions verbales figées contre 8000 ou 12000 verbes libres ; 6000 adverbes figés contre 2000 libres ; 80 000 noms simples contre 300 000 ou 400 000 noms composés) et leur forte probabilité d'occurrences dans les textes. Quand il y a ambiguïté potentielle, c'est presque toujours l'interprétation idiomatique qui est la bonne, sauf en cas de jeux de mots.